

A direct Proof for Quadratic Convergence of the Geometric Newton Method *

Markus Sprecher [†]

July 14, 2016

Introduction

We consider the problem of numerically computing a critical point of a functional $J: M \rightarrow \mathbb{R}$ where M is a Riemannian manifold. Due to local quadratic convergence a popular choice to solve this problem is the geometric Newton method. The proofs for quadratic convergence either use computations in a chart (Theorem 6.3.2 and Section 6.3.1 in [1]) or require additional geometric quantities such as parallel translation [3]. In this short note we provide a direct proof for quadratic convergence.

The Riemannian gradient $\text{grad}J(p) \in T_pM$ and Riemannian Hessian $\text{Hess}J(p): T_pM \rightarrow T_pM$ where T_pM is the tangent space at $p \in M$ are usually introduced using covariant derivatives. By Proposition 5.5.5 of [1], the Riemannian gradient and Riemannian Hessian can also be seen as the classical gradient and classical Hessian of the composition $J \circ R_p$ at 0, where $R_p: T_pM \rightarrow M$ is a retraction of order 2 as defined below.

Definition 1. Let M be a Riemannian manifold and $p \in M$. A C^∞ -map $R_p: T_pM \rightarrow M$ is called a retraction of order $k \in \mathbb{N}$ if

$$d(R_p(v), \exp_p(v)) \lesssim |v|^{k+1},$$

where $d: M \times M \rightarrow \mathbb{R}_{\geq 0}$ denotes the geodesic distance on M , $\exp_p: T_pM \rightarrow M$ the exponential map, $|\cdot|$ the norm induced by the inner product on T_pM and \lesssim indicates that there is a constant such that the left hand side can be estimated by a constant times the right hand side for $|v|$ small enough.

Taylor expansion of $J \circ R_p$ at 0 yields

$$J(R_p(v)) = J(p) + \langle v, \text{grad}J(p) \rangle + \frac{1}{2} \langle v, \text{Hess}J(p)v \rangle + \mathcal{O}(|v|^3), \quad (1)$$

where $\langle \cdot, \cdot \rangle$ is the inner product on T_pM and $\mathcal{O}(|v|^3)$ a term which can be estimated by a constant (depending on the third derivative of $J \circ R_p$ at 0) times $|v|^3$ for $|v|$ small enough.

Equation (1) together with the fact that $\text{Hess}J(p)$ is self-adjoint uniquely determines $\text{grad}J(p) \in T_pM$ and $\text{Hess}J(p): T_pM \rightarrow T_pM$ and we can therefore also use (1) as a definition for $\text{grad}J(p)$ and $\text{Hess}J(p)$.

Definition 2. Let M be a Riemannian manifold and R_p a retraction of order 1 at $p \in M$. Then the Riemannian gradient $\text{grad}J(p) \in T_pM$ and Riemannian Hessian $\text{Hess}J(p): T_pM \rightarrow T_pM$ are the unique vector respectively unique self-adjoint operator such that (1) holds.

*The research was supported by the Swiss National Fund under grant SNF 140635.

[†]Seminar for Applied Mathematics, ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland.

Note that here we require the retraction only to be of order 1. The Riemannian Hessian will in general depend on the choice of the retraction. However, if the retraction is of order 2 the Riemannian Hessian will coincide with the Riemannian Hessian of [1]. Definition 2 will allow us to give a direct proof for quadratic convergence. We can now define the geometric Newton method.

Definition 3. Let M be a Riemannian manifold, R_p a retraction of order 1 at $p \in M$ and $\text{grad}J(p) \in T_pM$ and $\text{Hess}J(p): T_pM \rightarrow T_pM$ as in Definition 2. Assume that $\text{Hess}J(p)$ is invertible. Then the geometric Newton method at $p \in M$ is given by the iteration

$$\phi(p) := R_p \left(-(\text{Hess}J(p))^{-1} \text{grad}J(p) \right). \quad (2)$$

Proof of quadratic convergence

We will need two lemmas to prove local quadratic convergence of the geometric Newton method. First, we show an estimate involving the Riemannian gradient and Riemannian Hessian. It can be regarded as a weak version of the fact that the derivative of the gradient is the Hessian.

Lemma 1. Let M be a Riemannian manifold, p^* a critical point of a three times differentiable functional $J: M \rightarrow \mathbb{R}$, $p \in M$, $R_p: T_pM \rightarrow M$ a retraction of order 1, $\text{grad}J(p) \in T_pM$ and $\text{Hess}J(p): T_pM \rightarrow T_pM$ as in Definition 2 and $v \in T_pM$ such that $p^* = R_p(v)$. Then we have

$$|\text{grad}J(p) + \text{Hess}J(p)v| \lesssim |v|^2. \quad (3)$$

The implicit constant depends only on the third derivative of $J \circ R_p$ at 0.

Proof. Let $w := \text{grad}J(p) + \text{Hess}J(p)v$. Replacing v in (1) by $v + tw$ and using that $\text{Hess}J(p)$ is self-adjoint yields

$$\begin{aligned} J(R_p(v + tw)) &= J(p) + \langle v + tw, \text{grad}J(p) \rangle + \frac{1}{2} \langle v + tw, \text{Hess}J(p)(v + tw) \rangle + \mathcal{O}(|v + tw|^3) \\ &= J(p) + \langle v, \text{grad}J(p) \rangle + \frac{1}{2} \langle v, \text{Hess}J(p)v \rangle + \mathcal{O}(|v|^3) \\ &\quad + t \left(\langle w, \text{grad}J(p) + \text{Hess}J(p)v \rangle + \mathcal{O}(|w||v|^2) \right) + \mathcal{O}(t^2) \\ &= J(p^*) + t \left(|w|^2 + \mathcal{O}(|w||v|^2) \right) + \mathcal{O}(t^2). \end{aligned}$$

Since p^* is critical point of J we have $J(R_p(v + tw)) = J(p^*) + \mathcal{O}(t^2)$ hence $|w|^2 = \mathcal{O}(|w||v|^2)$ and therefore $|w| \lesssim |v|^2$, which is equivalent to (3). \square

Additionally, we will need the following Lemma.

Lemma 2. Let M be a Riemannian manifold and R_p a retraction of order 1 at $p \in M$. Then we have

$$d(R_p(v), R_p(w)) = |v - w| + \mathcal{O}(|v|^2 + |w|^2).$$

Proof. By Nash embedding theorem [2] we can assume that $M \subset \mathbb{R}^n$ for some $n \in \mathbb{N}$. As R_p is a retraction of order 1 we have $R_p(v) = p + v + \mathcal{O}(|v|^2)$. Furthermore we have $d(p, q) = |p - q| + \mathcal{O}(|p - q|^2)$. Hence we have

$$d(R_p(v), R_p(w)) = |R_p(v) - R_p(w)| + \mathcal{O}(|R_p(v) - R_p(w)|^2) = |v - w| + \mathcal{O}(|v|^2 + |w|^2). \quad \square$$

We can now prove quadratic convergence.

Theorem 3. *Let M be a Riemannian manifold, p^* a critical point of a three times differentiable functional $J: M \rightarrow \mathbb{R}$, $p \in M$ and $\phi(p)$ as defined in (2). Assume that $\text{Hess}J(p^*)$ defined in Definition 2 is invertible. Then we have*

$$d(\phi(p), p^*) \lesssim d^2(p, p^*)$$

for all p in a neighborhood of p^ . The implicit constant depends only on the third derivative of $J \circ R_{p^*}$ at 0 and on the operator norm of $(\text{Hess}J(p^*))^{-1}$.*

Proof. Let v be as in Lemma 1. By differentiability of J and Lemma 1 we have

$$\begin{aligned} |(\text{Hess}J(p))^{-1} \text{grad}J(p) + v| &\lesssim \|(\text{Hess}J(p))^{-1}\|_{T_p M \rightarrow T_p M} |\text{grad}J(p) + \text{Hess}J(p)v| \\ &\lesssim \left(\|(\text{Hess}J(p^*))^{-1}\|_{T_p M \rightarrow T_p M} + \mathcal{O}(d(p, p^*)) \right) |v|^2 \\ &\lesssim |v|^2. \end{aligned}$$

By Lemma 2 we now have

$$d(\phi(p), p^*) = d(R_p(-(\text{Hess}J(p))^{-1} \text{grad}J(p)), R_p(v)) \lesssim |(\text{Hess}J(p))^{-1} \text{grad}J(p) + v| \lesssim |v|^2 \lesssim d^2(p, p^*). \quad \square$$

References

- [1] P.-A. Absil, R. Mahony, and R. Sepulchre. Optimization algorithms on matrix manifolds. Princeton University Press, 2008.
- [2] J. Nash. The imbedding problem for Riemannian manifolds. Annals of Mathematics, 63(1):20–63, 1956.
- [3] S. T. Smith. Optimization techniques on Riemannian manifolds. Fields Institute Communication, 3: 113–136, 1994.